Research Proposal

Yuqing (Candice) Zhang

Background

- Few studies have evaluated why restaurants failed
- Location, affiliation, and size are significant influences on restaurants' mortality[2]
  - But, to what extent?
  - Cofounding variables?

Research Question

- How do the mechanisms of food, ambience and location impact the **closing** of restaurants, which got their license in 2012 in Chicago area?

Background-Yelp Reviews

- Yelp is currently the most popular online consumer review website used for local business reviews and recommendations
- By the end of Q2 2016, yelpers have written more than 108 million reviews.
- 82 percent say their purchase decisions have been directly influenced by online reviews
- A one star increase in Yelp rating leads to a 5-9 percent increase in revenue[1]

Data

- YELP
  - Business
  - Reviews
- Business License

Business Data

yelp_academic_dataset_business.json

```
{
    "business_id":"encrypted business id",
    "name":"business name",
    "neighborhood":"hood name",
    "address":"full address",
    "city":"city",
    "state":"state -- if applicable --",
    "postal code":"postal code",
    "latitude":latitude,
    "longitude":longitude,
    "stars":star rating, rounded to half-stars,
    "review_count":number of reviews,
    "is_open":0/1 (closed/open),
    "attributes":["an array of strings: each array element is an attribute"],
    "categories":["an array of strings of business categories"],
    "hours":["an array of strings of business hours"],
    "type": "business"
}
```

Reviews

yelp_academic_dataset_review.json

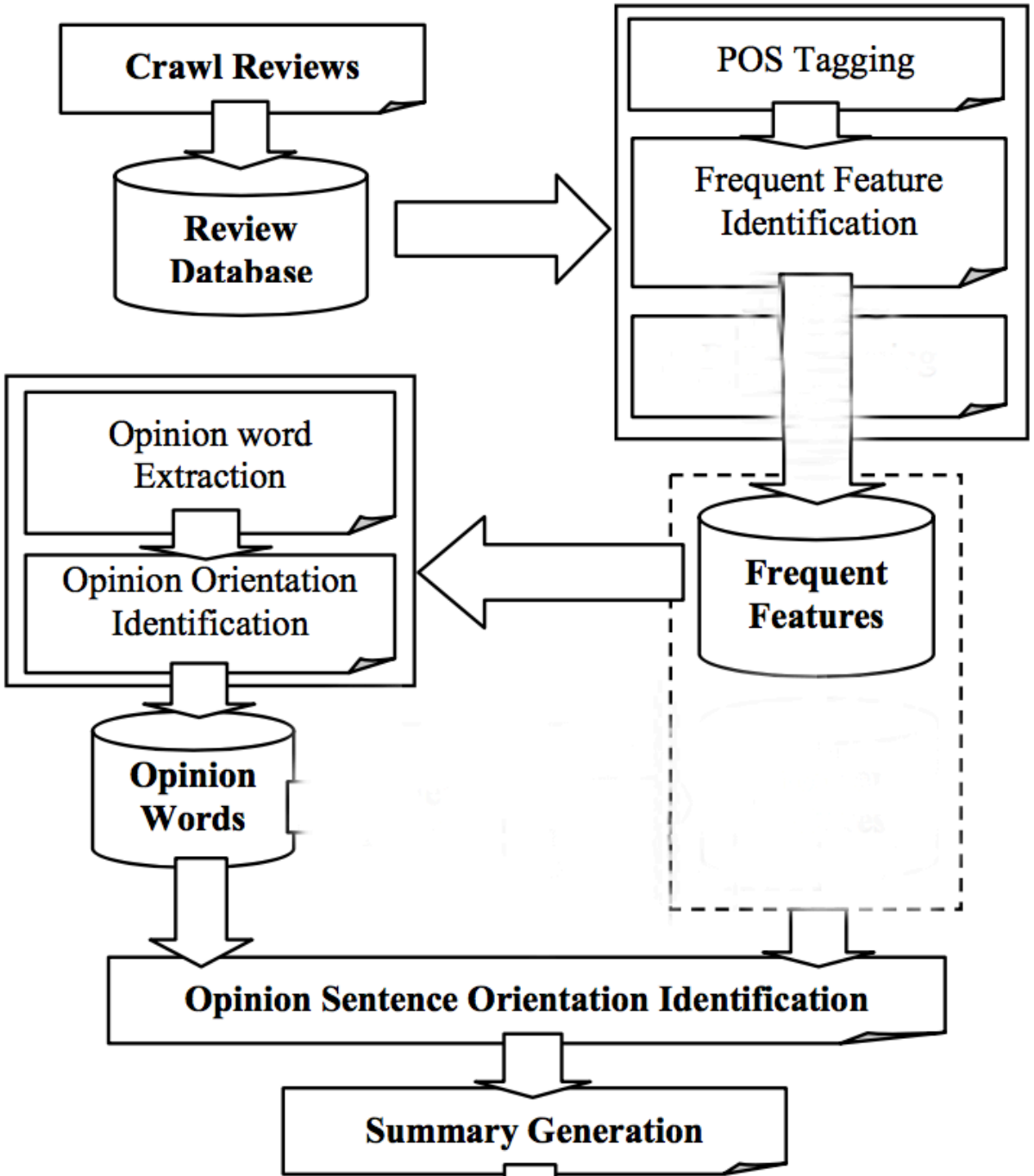
```
{
    "review_id":"encrypted review id",
    "user_id":"encrypted user id",
    "business_id":"encrypted business id",
    "stars":star rating, rounded to half-stars,
    "date":"date formatted like 2009-12-19",
    "text":"review text",
    "useful":number of useful votes received,
    "funny":number of funny votes received,
    "cool": number of cool review votes received,
    "type": "review"
}
```

License Issue Date

LEGAL NAME	DATE ISSUED
BELL OIL TERMINAL INC	8/11/06
BUCCI BIG & TALL INC.	8/30/16
PROJECT: VISION , INC.	6/22/16
FOLASHADE'S CLEANING SERVICE INC.	4/1/16
WALGREEN CO.	5/11/07
BURKS HEATING AND COOLING SOLUTIONS, LLC	8/30/16
BELL OIL TERMINAL INC	4/16/04
JAM PRODUCTIONS, LTD.	8/30/16
ANGELINE R. MC CARTHY	8/30/16
REVOLUTION BREWING, LLC	8/30/16
BELL OIL TERMINAL INC	3/5/04
WALGREEN CO.	4/28/03
	8/30/16
	8/30/16
	8/30/16

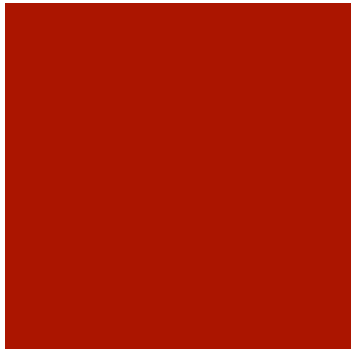
Methodology - Filter

- Include only
  - Restaurants
  - Chicago area
  - Got their license issued at 2012
  - Is closed
  - Split reviews into year 2012,2013,2014,2015,2016,2017



The Proposed Techniques

Methodology-Frequent Feature Identification



- Category Prediction
  - Trains on review data and generates a simple naïve-Bayes model that can predict the category of some text

`category_predictor` : Given some text, predict likely categories. For example:

```
$ python category_predictor/category_predictor.py yelp_academic_dataset.json > category_predictor.json
$ python category_predictor/predict.py category_predictor.json "bacon donut"
Category: "Food" - 82.66% chance
Category: "Restaurants" - 16.99% chance
Category: "Donuts" - 0.12% chance
Category: "Basque" - 0.02% chance
Category: "Spanish" - 0.02% chance
```

Methodology-Sentiment Analysis

- Natural Language Tool Kit
- **POS Tagging:** NLProcessor linguistic parser to parse each review to split text into sentences and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc)

- **FFI**: Food, location, ambience. Category Prediction Function from Yelp

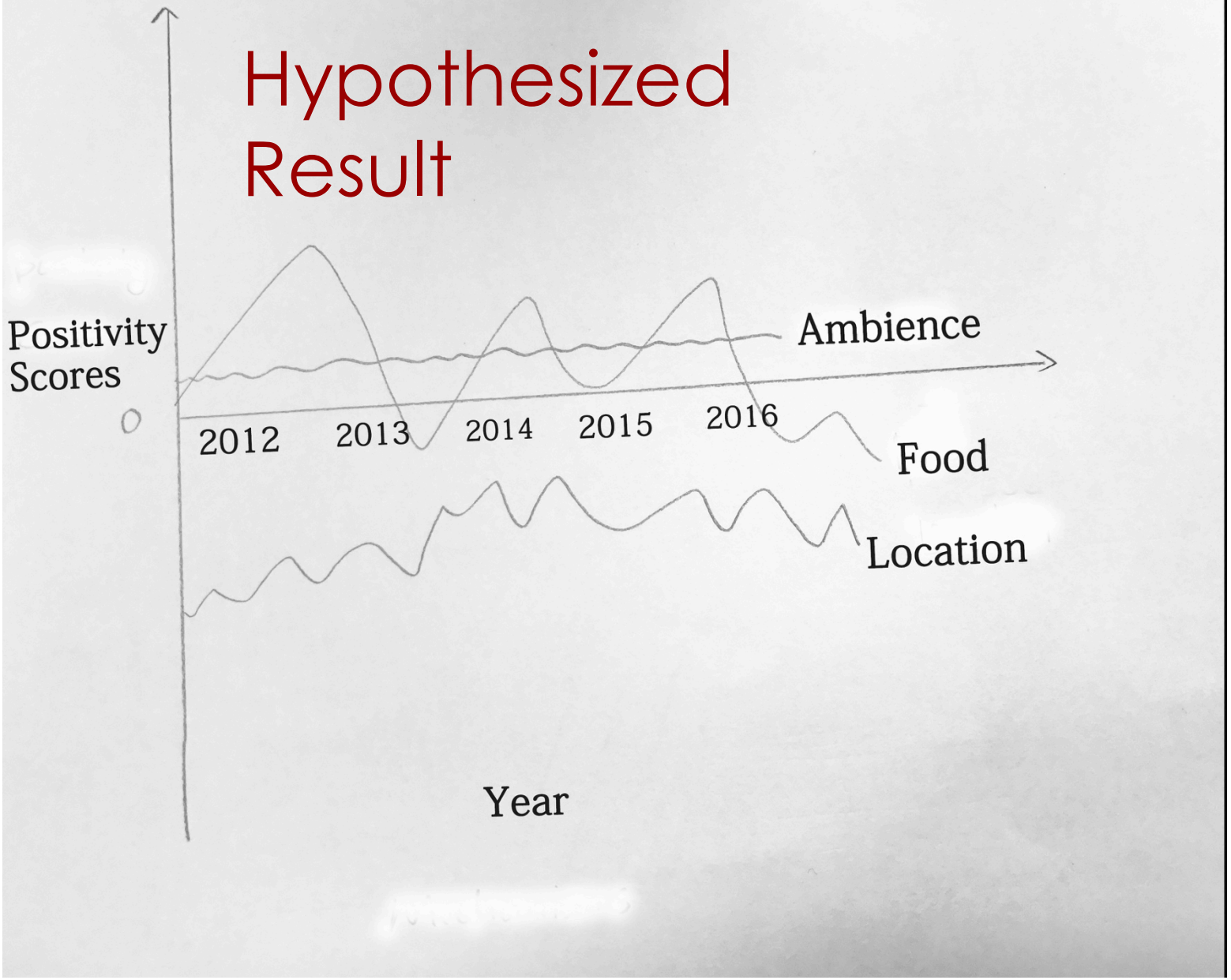
- **Opinion Words Extraction:**
  - Adjectives is useful for predicting whether a sentence is subjective, i.e., expressing an opinion.
  - Limit the opinion words extraction to those sentences that contain one or more product features

Methodology-Sentiment Analysis

- **Orientation Identification for Opinion Words**:
  - SentimentIntensityAnalyzer function from NLTK to calculate a positivity score for each word

- **Predicting the Orientations of Opinion Sentences**:
  - Use the dominant orientation of the opinion words in the sentence to determine the orientation of the sentence

Hypothesized Result



Reference

- [1]Luca, Michael. "Reviews, Reputation, and Revenue: The Case of Yelp.com.' Harvard Business School Working Paper, No. 12-016, September 2011. (Revised March 2016. Revise and resubmit at the *American Economic Journal – Applied Economics*.)

- [2]Parsa, H. G., Self, J., Sydnor-Busso, S., & Yoon, H. J. (2011). Why Restaurants Fail? Part II - The Impact of Affiliation, Location, and Size on Restaurant Failures: Results from a Survival Analysis. Journal of Foodservice Business Research, 14(4), 360-379. doi: 10.1080/15378020.2011.625824